

Article court Estimation de mélanges généralisés dans les arbres de Markov cachés, application à la segmentation des images de cartons d'orgue de barbarie

Short paper Generalized mixture estimation in hidden Markov trees,
application to segmentation of images of street organ cards

Emmanuel Monfrini * et Wojciech Pieczynski **

*Université Paris VI, LSTA, 175 rue du Chevaleret, 75013 Paris, France

**GET/INT Département CITI, CNRS UMR 5157, 9, rue Charles Fourier, 91000 Evry, France
tél. 01.60.76.44.25, fax 01.60.76.44.33, e-mail : Wojciech.Pieczynski@int-evry.fr

Manuscrit reçu le 30 octobre 2003

Résumé et mots clés

Nous nous intéressons dans cet article à la segmentation statistique non supervisée d'images avec les modèles par arbres de Markov cachés. Dans un premier temps nous proposons deux modèles évolutifs originaux et étudions, via simulations, le comportement des diverses méthodes générales de l'estimation des paramètres. Ensuite, nous proposons une méthode originale d'estimation de mélanges généralisés et montrons son bon comportement, même dans des cas d'images très fortement bruitées, par une étude de simulations. Cette même méthode est appliquée au problème de la segmentation des cartons d'orgue de barbarie, attestant de son intérêt dans une situation réelle. Arbres de Markov cachés, segmentation statistique d'images, mélange généralisé, EM, ICE, SEM.

Abstract and key words

We deal in this paper with unsupervised statistical image segmentation using hidden Markov trees. First, we propose two original evolutionary models and study, via simulations, the behaviour of different general estimation methods. Second, we propose a new generalized mixture estimation method and show its efficiency in unsupervised image segmentation, even in very noisy settings. The proposed method is then successfully applied to the unsupervised segmentation of street organ cards images. Hidden Markov trees, statistical image segmentation, generalized mixture, EM, ICE, SEM.

1. Introduction

La problématique de la segmentation statistique d'images considérée dans ce travail est la suivante : on considère deux processus aléatoires $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$, où S est l'ensemble des pixels. Chaque X_s est à valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et chaque Y_s est à valeurs dans \mathbb{R} . Il s'agit d'estimer la réalisation inobservable x de X à partir de l'observation y de Y . Pour cela on choisit généralement de modéliser le processus caché X sous forme markovienne (champs, chaînes ou arbres de Markov) et de restreindre la famille des lois de Y conditionnelle à X , qui modélisent le bruit au sens large, de façon à ce que la loi de X conditionnelle à Y soit encore markovienne. Il devient alors possible d'appliquer les méthodes bayésiennes d'estimation Maximum a Posteriori (MPM) ou Maximum a Posteriori (MAP). De plus, les paramètres peuvent très souvent être estimés par des méthodes générales de type « Expectation-

Maximization » (EM [14]) ou « Iterative Conditional Estimation (ICE [8-10]), ce qui permet la conception des méthodes de segmentation non supervisées.

Nous proposons dans cet article une étude de segmentation non supervisée d'images par le modèle des arbres de Markov cachés [2, 4-7, 11-13, 17, 19, 20], en mettant l'accent sur le problème de l'estimation des paramètres. Pour cela, nous avons, dans un premier temps évalué et comparé différents algorithmes d'estimation des paramètres dans le cas gaussien. Les résultats numériques correspondant à cette étude sont proposés dans la section 2. Dans un second temps, nous avons étudié l'extension de la modélisation gaussienne classique aux cas où la forme des densités de bruit ne sont pas connues a priori et peuvent varier avec la classe ω_i . Cette extension de la modélisation du bruit, introduite dans les chaînes et champs de Markov cachés respectivement dans [10] et [8], est essentiellement motivée par le souhait de disposer d'une modélisation de nature à pouvoir mieux prendre en compte la diversité des textures naturelles.

L'exemple de la lecture de cartons d'orgue de barbarie, étudié dans la section 3, montre l'utilité de ce type de modélisation, notamment lorsqu'il s'agit de lire des cartons anciens. Les conclusions et perspectives constituent la dernière section.

2. Arbres de Markov cachés

2.1. Modélisations sur arbres

Nous considérons le cas des images mono résolution, qui correspond aux arbres dont les « feuilles » indiquent les variables du processus observé. On « regroupe » quatre à quatre ou deux à deux les variables du processus et on construit un père à chaque groupe de feuilles. Cette construction est alors appliquée au nouveau processus formé par les pères. On répète ainsi de suite l'opération tant qu'elle est possible jusqu'à l'ancêtre unique, que l'on appelle la « racine ». Formellement, soit $X = (X_s)_{s \in S}$, le processus des classes et $Y = (Y_s)_{s \in S}$ la variable observable. Soit S^1, \dots, S^N une partition de S représentant les différentes « générations », qui seront appelées « échelles » dans la suite, du processus sur arbre. À chaque $s \in S^i$ correspond $s^+ \subset S^{i+1}$ appelé l'ensemble des « fils » de s , les ensembles des enfants des éléments de S^i forment une partition de S^{i+1} . Par ailleurs, on suppose que S^1 admet un seul élément s_r appelé « racine », on note par s^{++} l'ensemble de tous les descendants de s , et par s^- son unique « père ».

Le quadarbre est un arbre pour lequel chaque père a quatre « fils ». Sa structure est bien adaptée aux images qui font couramment $2^N \times 2^N$ pixels. Le diarbre, pour lequel chaque père a deux « fils », génère, une échelle sur deux, un feuillage rectangulaire. Cependant, la construction d'un diarbre, à partir d'une image, se fait suivant le même schéma que celle d'un quadarbre. Le diarbre est un modèle strictement plus général que le quadarbre ; en effet, ce dernier peut être vu comme un diarbre particulier dans lequel les pères sont, de manière déterministe, identiques au grand-père. Par ailleurs, une latitude supplémentaire s'offre, lors de la construction du diarbre, au niveau du choix du regroupement des fils. En effet, il y a trois façons de

regrouper deux à deux quatre petits-fils [15]. Nous ne considérerons ici que le cas où le lien de fraternité est basé sur le voisinage droit. Les fils sont engendrés alternativement dans le sens latéral et dans la profondeur, comme indiqué à la Figure 1. Une étude concernant les autres diarbres et leur comparaison avec les quadarbres, qui désavantage légèrement ces derniers, peut être consultée dans [15].

1.2. Les arbres évolutifs

Le processus des classes $X = (X_s)_{s \in S}$, chaque X_s prenant ses valeurs dans $\Omega = \{\omega_1, \dots, \omega_K\}$, est un arbre de Markov si sa loi s'écrit

$$p(x) = p(x_{s_r}) \prod_{s \in S - S^1} p(x_s | x_{s^-}) \quad (1)$$

Le modèle d'arbre le plus simple est celui pour lequel les probabilités de transition père-fils associées au couple de réalisations (ω_i, ω_j) sont constantes, quelle que soit l'échelle. Ce type de modèle peut cependant présenter des faiblesses ; en effet, contrairement à ce qui se passe avec les champs de Markov, on peut rencontrer des difficultés à obtenir des zones d'uniformité des classes. Pour remédier à ce problème, il est possible de complexifier le modèle en faisant dépendre les probabilités de transition de l'échelle [2, 20]. Nous reprenons cette idée en augmentant la probabilité de conservation des classes au fur et à mesure de l'affinage de la résolution, ce qui devra, au plan intuitif, désavantager l'apparition des points isolés et des « petites » taches. Dans les deux modélisations retenues les probabilités de transitions inter-échelles dépendent de deux paramètres α, ε , et de la résolution. Parmi les différentes formes de fonctions faisant intervenir la résolution testées, nous en avons retenu deux, qui donnent les résultats les plus intéressants, particulièrement en ce qui concerne la concentration des zones de classes uniformes. En

posant $A(n) = \frac{\ln(N) - \ln(n)}{\ln(N)}$ et $B(n) = \sqrt{\frac{N-n}{N}}$, nous considérons les deux types des probabilités de transition père-fils suivantes :

$$\begin{aligned} g_{i,j}^n(\alpha) &= p(x_s^n = \omega_i | x_{s^-}^{n-1} = \omega_j) \\ &= \begin{cases} 1 - \varepsilon - \alpha A(n) & \text{pour } i = j \\ (\varepsilon + \alpha A(n)) / (K - 1) & \text{pour } i \neq j \end{cases} \end{aligned} \quad (2)$$

$$\begin{aligned} h_{i,j}^n(\alpha) &= p(x_s^n = \omega_i | x_{s^-}^{n-1} = \omega_j) \\ &= \begin{cases} 1 - \varepsilon - \alpha B(n) & \text{pour } i = j \\ (\varepsilon + \alpha B(n)) / (K - 1) & \text{pour } i \neq j \end{cases} \end{aligned} \quad (3)$$

Le paramètre ε permet de fixer les transitions de l'avant-dernière génération vers la génération la plus fine. Un tel modèle des arbres de Markov, qui sera dit « évolutif » dans la suite, peut améliorer sensiblement la qualité des segmentations comme le montrent les différents résultats présentés dans [15].

Dans la suite nous qualifierons de « type 1 » les développements effectués dans le cadre des probabilités de transition (2) et de « type 2 » ceux issus des probabilités de transition (3).

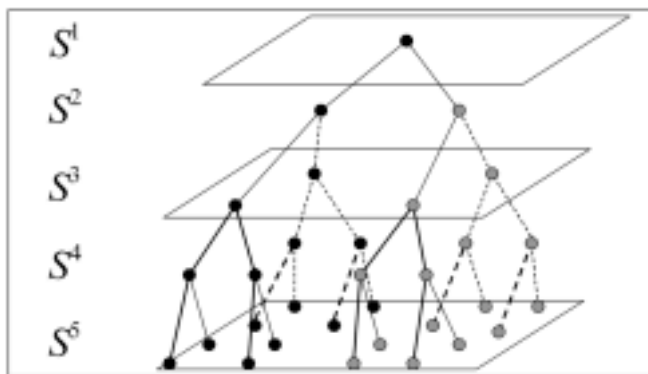


Figure 1. Diarbre droit.

2.3. Modèle markovien caché sur arbre

En reprenant la notation $X = (X_s)_{s \in S}$, $Y = (Y_s)_{s \in S}$, la loi du couple (X, Y) s'écrit $p(x, y) = p(x)p(y|x)$. Nous considérons dans la suite le modèle d'Arbre de Markov caché (AMC) classique dans lequel $p(x)$ est donnée par (1), et $p(y|x)$ est de la forme $p(y|x) = \prod_{s \in S} p(y_s|x_s)$ (avec $p(y_s|x_s) = 1$ pour $s \notin S^N$).

On montre alors que la loi de X a posteriori $p(x|y)$ admet la même structure markovienne que sa loi a priori $p(x)$, avec les transitions $p(x_s|x_{s-}, y)$ et les marginales $p(x_s|y)$ calculables [13]. Plus précisément, les probabilités « retrogrades » $\beta(x_s) = p(y_{s++}|z_s)$, qui peuvent être calculées récursivement par $\beta(x_s) = 1$ pour $s \in S^N$, et $\beta(x_s) = \prod_{t \in S^+} (\sum_{x_t \in \Omega} \beta(x_t)p(z_t|z_s))$ pour $s \in S - S^N$, déterminent

$$p(x_s|x_{s-}, y) \text{ par } p^y(x_s|x_{s-}) = \frac{\beta(x_s)p(z_s|z_{s-})}{\sum_{\omega_s \in \Omega} \beta(\omega_s)p(\omega_s, y_s|z_{s-})}.$$

Toutes les transitions $p(x_s|x_{s-}, y)$ ayant été calculées dans une « passe ascendante » utilisant les récursions ci-dessus, on peut en déduire les marginales a posteriori $p(x_s|y)$. En effet, pour $s \in S - S^1$ quelconque il existe un chemin unique $c = (s_1, \dots, s_{i-1}, s)$ le reliant au sommet s_1 , et la restriction de X à ce chemin est, pour sa loi conditionnelle à $Y = y$, une chaîne de Markov. $p(x_s|y)$ se calcule alors classiquement de proche en proche par $p^y(x_{s_j}) = \sum_{x_{s_{j-1}} \in \Omega} p^y(x_{s_{j-1}})p^y(x_{s_j}|x_{s_{j-1}})$. Cela per-

met la simulation de X selon sa loi a posteriori, importante dans la problématique de l'estimation des paramètres abordée dans la section suivante. Par ailleurs, la méthode de segmentation MPM est obtenue en maximisant $p(x_s|y)$ pour tout $s \in S^N$.

3. Estimation des paramètres dans les modèles cachés classiques

Nous cherchons à estimer les paramètres $\theta = \{\theta_1, \theta_2\}$ où θ_1 est l'ensemble des paramètres concernant la loi de X et θ_2 l'ensemble de ceux concernant les lois de Y conditionnellement à X , que l'on supposera gaussiennes, dans cette section. L'estimation se fait à partir de la seule observation des réalisations de $(Y_s)_{s \in S^N}$. La loi de X est donnée par les probabilités de la racine sur chaque classe $\pi_i = p(x_{s_r} = \omega_i)$, et les probabilités de transition père-fils aux différentes échelles, qui sont données par (2) ou (3). Le paramètre ε étant fixé à 0.001 dans la suite, ces transitions sont entièrement définies par la connaissance du paramètre α . Les lois de Y conditionnellement à X sont définies par les paramètres du bruit, que sont les moyennes et variances des distributions normales. On les notera, pour tout $1 \leq i \leq K$, respectivement μ_i et σ_i^2 . Ainsi dans les deux cas considérés les paramètres $\theta = (\theta_1, \theta_2)$ sont $\theta_1 = (\pi_1, \dots, \pi_K, \alpha)$, et $\theta_2 = (\mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2)$. Nous considérons trois méthodes itératives de leur estimation, chacune déterminant une suite $\theta^1, \theta^2, \dots, \theta^q, \dots$. Après avoir initialisé les différentes méthodes par un même algorithme, nous précisons le calcul, à l'itération q , de

θ^{q+1} à partir de θ^q et $Y = y$. L'initialisation commune aux trois méthodes est la suivante. Les π_i sont toutes initialisées par $\pi_i^1 = 1/K$, et α est initialisé par $\alpha^1 = 1$. Les variances sont toutes initialisées par la variance empirique de l'échantillon formé par l'ensemble des observations $(y_s)_{s \in S^N}$, et pour tout $1 \leq i \leq K$, la moyenne μ_i est initialisée par $\mu_i^1 = 2Mi/(K+1)$, où M est la moyenne des observations $(y_s)_{s \in S^N}$. La simplicité relative de cette initialisation montre accessoirement le bon comportement des trois méthodes dans le cadre des expérimentations présentées ci-dessous.

3.1. Algorithme EM sur le diarbre

L'algorithme « Expectation-Maximization » (EM) est parmi les plus utilisés dans les problèmes de l'estimation des paramètres dans le cas des données cachées [14]. En particulier, il donne de bons résultats dans le cas des AMC [11-13]. Nous proposons une méthode « de type EM », qui peut être vue comme une « généralisation » du EM au cas évolutif considéré. Pour la ré-estimation de α , on reprend les formules de ré-estimations du EM à la racine et à chaque échelle. Ces formules donnent les probabilités de transitions, qui permettent le calcul de α , sur chaque échelle, en inversant g^n (ou h^n , selon le modèle utilisé). Ensuite, la valeur suivante α^{q+1} de α est donnée par la moyenne, pondérée par les cardinaux des échantillons sur chaque échelle. Plus précisément, nous avons :

$$\pi_i^{q+1} = p(x_{s-} = \omega_i | y, \theta^q) \quad (4)$$

et, pour tout n , $2 \leq n \leq N$,

$$g_{ij}^n(\alpha_n^{q+1}) = \frac{\sum_{s \in S^n} p(x_s^n = \omega_i | x_{s-}^{n-1} = \omega_j, y, \theta^q)}{|S^n|} \quad (5)$$

La valeur suivante α^{q+1} de α est donc calculée à partir des $N-1$ valeurs $\alpha_2^q, \dots, \alpha_N^q$ obtenues par (5), chacune d'elles étant obtenue en inversant g_{ij}^n (ou h_{ij}^n pour le modèle 2) correspondante, en prenant la moyenne pondérée par les cardinaux des ensembles S^2, \dots, S^N de ces valeurs :

$$\alpha^{(q+1)} = \frac{|S^2|\alpha_2^{q+1} + \dots + |S^N|\alpha_N^{q+1}}{|S^2| + \dots + |S^N|} \quad (6)$$

La mise à jour des paramètres du bruit θ_2 se fait classiquement par :

$$\mu_i^{q+1} = \frac{\sum_{s \in S^N} y_s p(x_s = \omega_i | y, \theta^q)}{\sum_{s \in S^N} p(x_s = \omega_i | y, \theta^q)} \quad (7)$$

$$(\sigma_i^2)^{q+1} = \frac{\sum_{s \in S^N} (y_s - \mu_i^{q+1})^2 p(x_s = \omega_i | y, \theta^q)}{\sum_{s \in S^N} p(x_s = \omega_i | y, \theta^q)} \quad (8)$$

3.3. Algorithmes ICE et MICE

Nous mettons en place, sur ce même modèle deux autres procédures d'estimation des paramètres, « Iterative conditional estimation » (ICE) et « ICE modifiée » (MICE). Selon le principe de l'ICE, on se donne un estimateur de $\theta = (\theta_1, \theta_2)$ à partir des données complètes $\hat{\theta}(X, Y)$, et les itérations sont définies par $\theta^{q+1} = E_{\theta^q}[\hat{\theta}(X, Y) | Y = y]$. Cette espérance peut être calculable pour certaines composantes de θ , auquel cas on dispose des formules analytiques, et ne pas l'être pour les composantes restantes, pour lesquelles elle est approchée en utilisant des tirages selon la loi de X a posteriori. Sachant que $\hat{\theta}(X, Y) = (\hat{\theta}_1(X), \hat{\theta}_2(X, Y))$, nous proposons d'estimer $g_{ij}^n(\alpha)$ à partir de (X^{n-1}, X^n) par la fréquence empirique. L'espérance de ce dernier estimateur est alors calculable et le résultat obtenu est identique à celui donné par la méthode EM [1, 9]. La ré-estimation de α et π est ainsi la même que dans la sous-section précédente. Concernant l'estimation à partir des données complètes des paramètres du bruit θ_2 , on utilise classiquement les

moyennes et les variances empiriques, que l'on notera $\hat{\theta}_2(X, Y)$. L'espérance $E_{\theta^{(q)}}[\hat{\theta}_2(X, Y) | Y = y]$ n'est pas calculable et nous l'approchons sur la base d'un seul tirage $X = x^q$ selon la loi de X conditionnelle à $Y = y$, ce dernier utilisant le paramètre courant θ^q . Ainsi $\theta_2^{q+1} = \hat{\theta}_2(x^q, y)$, ce qui est détaillé dans les formules (9) et (10).

$$\mu_i^{q+1} = \frac{\sum_{s \in S^N} y_s 1_{[x_s = \omega_i]}}{\sum_{s \in S^N} 1_{[x_s = \omega_i]}} \tag{9}$$

$$(\sigma_i^2)^{q+1} = \frac{\sum_{s \in S^N} (y_s - \mu_i^{q+1})^2 1_{[x_s = \omega_i]}}{\sum_{s \in S^N} 1_{[x_s = \omega_i]}} \tag{10}$$

L'idée de l'algorithme MICE, proposée dans [1] dans le cadre des chaînes de Markov cachées, consiste en une modification de la manière de simuler $X = x^q = (x_s^q)_{s \in S^N}$, le reste de la démarche étant identique à celle utilisée dans l'ICE. En effet,



Figure 2. Images des classes considérées.

Tableau 1. Segmentations non supervisées méthode des mélanges généralisés, cas N-N.

Image	SEM type 1	ICE type1	MICE type1	EM type 1	SEM type 2	ICE type2	MICE type2	EM type 2	
	3.41%	3.51%	3.43%	3.51%	3.42%	3.40%	3.42%	3.43%	Erreur mini.
	4.29%	4.63%	4.66%	4.48%	4.72%	4.50%	4.46%	4.64%	Erreur maxi.
	3.91%	4.01%	3.96%	3.97%	4.01%	3.97%	3.95%	3.97%	Erreur moy.
	4.12%	4.23%	4.26%	4.03%	4.09%	4.02%	4.02%	4.12%	Erreur mini.
	6.41%	7.50%	6.45%	6.61%	6.52%	6.68%	7.84%	6.73%	Erreur maxi.
	5.04%	5.48%	5.32%	5.23%	5.16%	5.32%	5.22%	5.34%	Erreur moy.
	7.66%	8.14%	8.23%	7.84%	7.71%	7.60%	7.87%	7.67%	Erreur mini.
	9.86%	11.16%	11.06%	10.08%	9.81%	10.16%	10.35%	10.17%	Erreur maxi.
	8.86%	9.44%	9.41%	8.95%	8.75%	8.88%	8.88%	8.94%	Erreur moy.
	13.50%	14.87%	14.41%	13.83%	13.26%	13.78%	13.79%	13.89%	Erreur mini.
	15.39%	16.91%	16.82%	15.64%	15.06%	15.77%	15.80%	15.71%	Erreur maxi.
	14.51%	15.79%	15.62%	14.88%	14.15%	14.79%	14.68%	14.96%	Erreur moy.
	29.75%	32.23%	31.99%	30.37%	27.03%	29.87%	29.31%	31.15%	Erreur mini.
	31.91%	51.46%	47.42%	33.10%	32.24%	35.05%	34.01%	48.82%	Erreur maxi.
	30.66%	34.06%	33.88%	31.47%	28.46%	31.61%	31.45%	32.98%	Erreur moy.
	12.60%	13.76%	13.64%	12.90%	12.11%	12.91%	12.83%	13.24%	Erreur globale

dans MICE chaque X_s est simulée selon sa loi conditionnelle à $Y = y$ (la même donc que celle utilisée, après l'étape de l'estimation des paramètres, dans la segmentation MPM).

3.3. Algorithme SEM sur le diarbre

L'algorithme « Stochastic EM » (SEM), présenté comme une approximation stochastique de l'algorithme EM [3], est une méthode itérative dans laquelle on considère l'estimateur à partir des données complètes $\hat{\theta}(X, Y) = (\hat{\theta}_1(X), \hat{\theta}_2(X, Y))$ et, à chaque itération, on simule $X = x^q$ selon la loi de X conditionnelle à $Y = y$, utilisant le paramètre courant θ^q . Le paramètre suivant est donné par $\theta^{(q+1)} = \hat{\theta}(x^q, y)$. Nous utilisons le même estimateur $\hat{\theta}(X, Y) = (\hat{\theta}_1(X), \hat{\theta}_2(X, Y))$ que dans la sous-section précédente.

3.4. Choix de l'algorithme d'estimation des paramètres

Nous sommes désormais en présence de quatre algorithmes d'estimation des paramètres, adaptés à deux modèles évolutifs de diarbres différenciés par la fonction définissant les probabilités de transition inter-échelles. Nous comparons l'efficacité des huit algorithmes MPM correspondant dans le cadre de la segmentation non supervisée de cinq images présentées sur la Figure 2. Les cinq images sont bruitées par le bruit gaussien $N(0, 1)$, $N(1, 1)$ et l'objectif de notre étude est de déterminer l'algorithme le mieux adapté à chacun des deux modèles évolutifs en présence. Notons que malgré le bruitage identique le niveau de bruit croît lorsque l'homogénéité des classes décroît (en allant de gauche à droite dans la Figure 2) ; ce phénomène, observé dans le cas des chaînes et champs de Markov cachés [18], est dû à la dépendance du niveau de bruit de la loi a priori. Le Tableau 1 récapitule, pour chaque image et chaque algorithme, l'erreur moyenne sur les cent segmentations, ainsi que l'erreur minimale et l'erreur maximale commises au cours des cent expériences. Nous y donnons également l'erreur globale commise par chaque algorithme, au cours des cinq cent segmentations.

Globalement, nous notons que la méthode SEM est la plus intéressante pour ce qui concerne la segmentation non supervisée des images utilisant les diarbres en présence. Par ailleurs, nous constatons que les diarbres de type 2 semblent, globalement, plus intéressants que les diarbres de type 1.

4. Le modèle des mélanges généralisés

4.1. Présentation de la méthode

Dans un modèle de mélanges généralisé la forme du bruit peut varier avec la classe. L'idée de ce type de modélisation vient de deux constats : (i) plusieurs formes différentes correspondant aux différentes classes peuvent exister dans une même image ; de plus, à une classe donnée la forme du bruit peut varier avec

le temps [8] ; (ii) dans les cas classiques, les segmentations supervisées et non supervisées sont assez proches jusqu'à un niveau élevé du bruit. Le premier point montre l'intérêt pratique de telles méthodes, et le deuxième montre qu'il existe une certaine marge de manœuvre pour complexifier le modèle. Bien entendu, tout dépend du niveau de bruit, ce dernier étant difficile à définir dans le cadre markovien car dépendant également de la loi a priori. Diverses méthodes d'estimation de tels « mélanges généralisés » ont été proposées dans le cadre des champs [8] ou chaînes [10] de Markov et les résultats obtenus se sont montrés encourageants. Décrivons, à titre d'exemple, une de ces méthodes dite SEM généralisé (SEMG), qui est effectivement une généralisation de l'algorithme SEM. Supposons que pour chaque classe ω_i , la densité $p(y_s | x_s = \omega_i)$ peut prendre l'une des j_i formes $F_1^i, \dots, F_{j_i}^i$ possibles (pas exemple, F_1^i sont des densités gaussiennes, F_2^i sont des densités gamma, F_3^i sont des densités beta, ...). Dans cette méthode itérative on utilise, à chaque itération q , une réalisation x^q de X simulée selon sa loi conditionnelle à $Y = y$. Pour chaque classe ω_i , et pour chaque forme $F_1^i, \dots, F_{j_i}^i$ du bruit admissible pour ω_i , on utilise alors x^q et $Y = y$ pour estimer les paramètres par un estimateur défini à partir des données complètes (x^q, y) ; ce qui fournit pour chaque classe ω_i un « candidat » – toujours à l'itération q donnée – par forme admissible. On dispose ainsi, pour chaque classe ω_i , de j_i candidats $f_1^{i,q} \in F_1^i, \dots, f_{j_i}^{i,q} \in F_{j_i}^i$. On choisit alors parmi les candidats $f_1^{i,q}, \dots, f_{j_i}^{i,q}$ celui qui correspond le mieux, selon un critère retenu qui peut être la minimisation d'une distance, à une densité estimée de manière non paramétrique comme, à titre d'exemple, l'histogramme. Le candidat retenu est ainsi la densité suivante $f^{i,q+1}$, qui peut être de la même forme que la densité précédente ou pas. Ainsi la forme – et les paramètres – du bruit peuvent varier au cours des itérations. Notons que de telles méthodes autorisent la considération d'un grand nombre des formes admissibles pour chacune des classes, sans augmentation rédhibitoire de temps de calcul par rapport aux méthodes de l'estimation de mélanges classiques. De plus, ces dernières apparaissent comme en étant des cas particuliers.

Nous proposons ici une méthode d'estimation de mélanges généralisés différente, qui est possible de par la rapidité des calculs dans le cadre des arbres de Markov. Il devient, en effet, envisageable de faire une estimation complète des paramètres pour chaque cas de mélange de lois classique autorisé par les familles de distributions considérées. Plus précisément, considérons K classes et supposons que pour chaque classe ω_i le bruit prend l'une des j_i formes $\{F_1^i, \dots, F_{j_i}^i\}$ admissibles. Nous avons alors $j_1 \dots j_K$ mélanges classiques possibles, chacun étant estimé par SEM 1 et SEM 2. Une fois les différentes estimations effectuées, on choisit parmi les $j_1 \dots j_K$ solutions celle qui décrit le mieux, selon un certain critère, le processus étudié. À titre d'exemple, on peut choisir la solution qui minimisera une distance entre la distribution théorique des observations donnée par le mélange estimé, et la distribution de l'échantillon, approchée par un histogramme, ou par un estimateur à noyau de la distribution de mélange. Nous proposons une méthode de choix parmi les $j_1 \dots j_K$ solutions fondée sur les moments. D'une part,

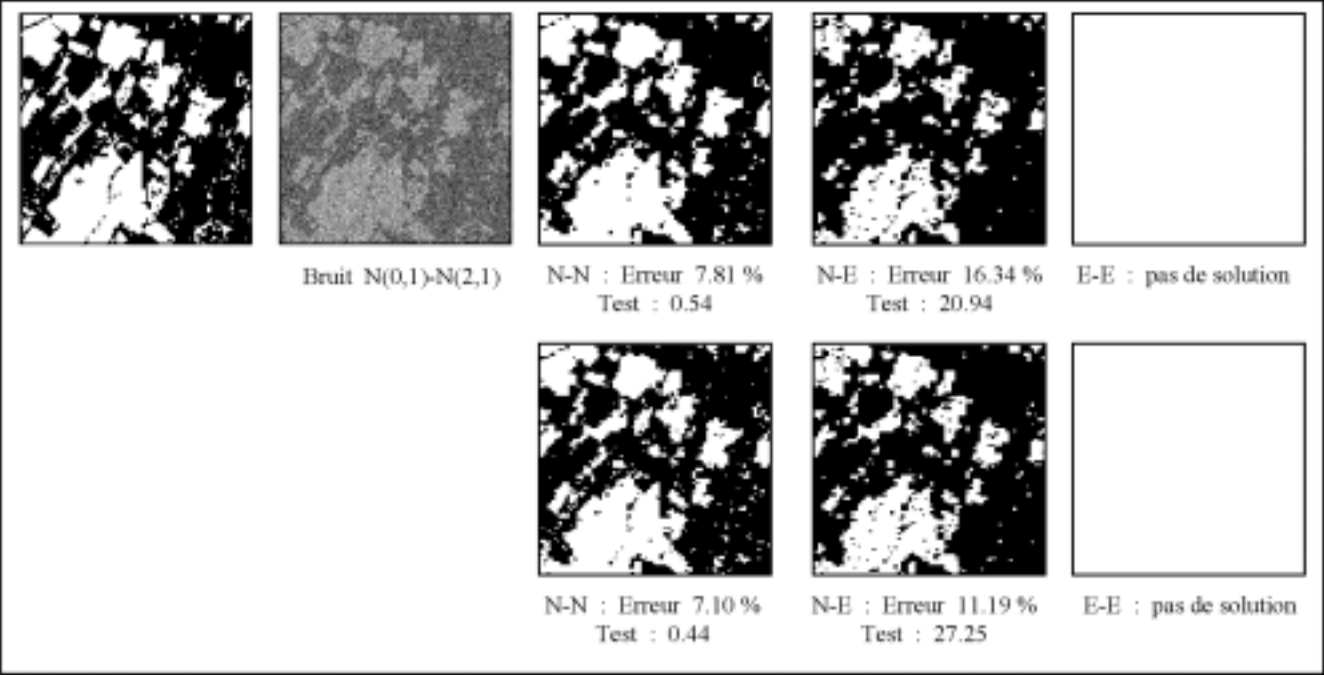


Figure 3. Segmentations non supervisées méthode des mélanges généralisés.

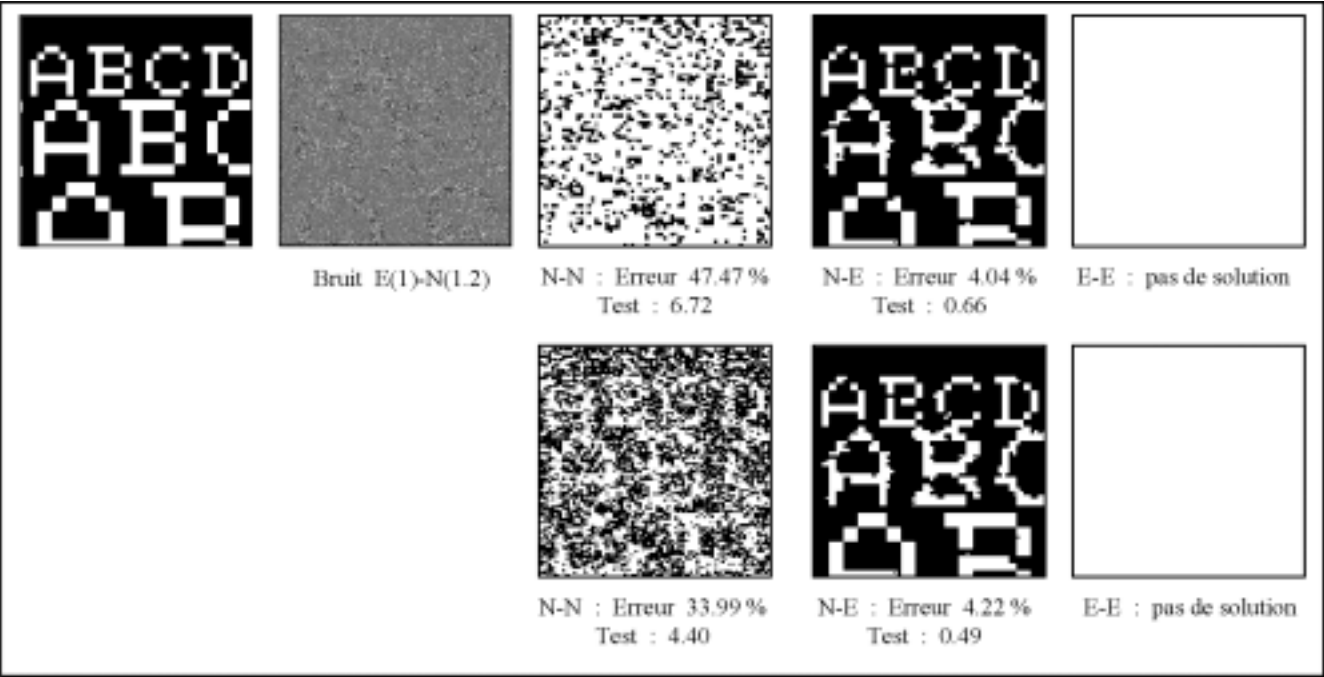


Figure 4. Segmentations non supervisées méthode des mélanges généralisés.

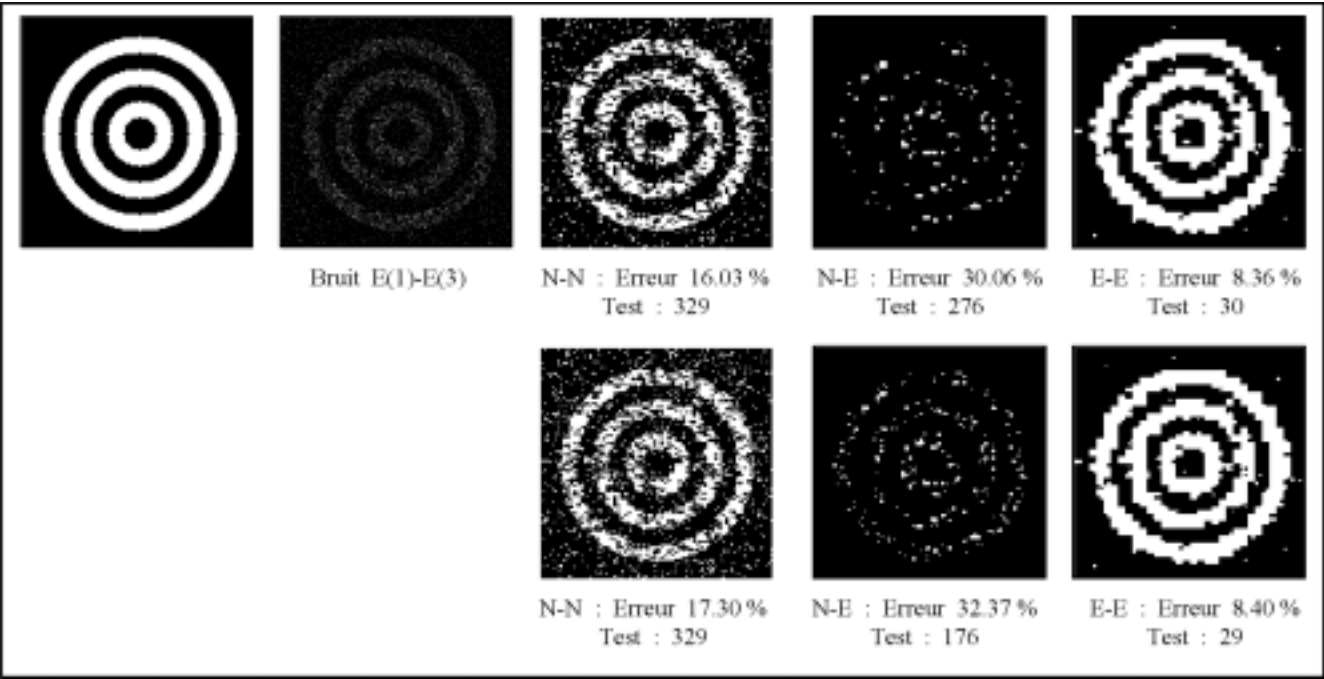


Figure 5. Segmentations non supervisées méthode des mélanges généralisés.

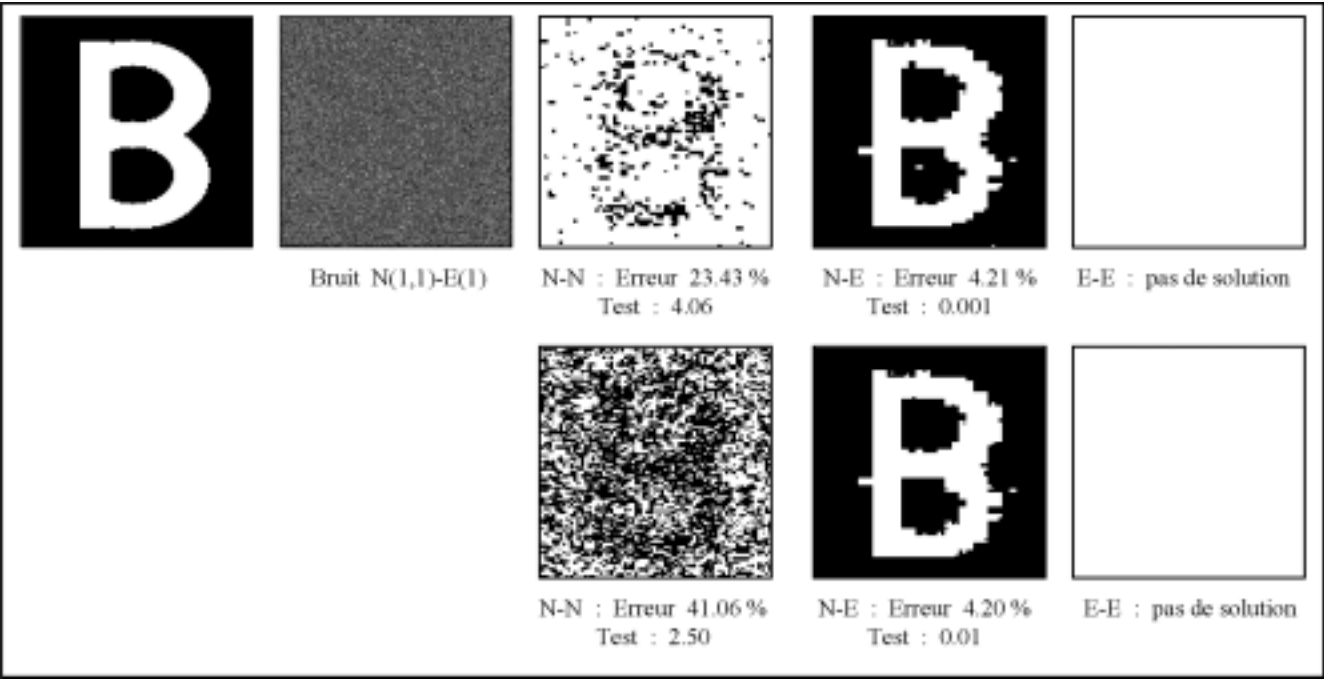


Figure 6. Segmentations non supervisées méthode des mélanges généralisés.

on calcule les quatre premiers moments empiriques $\hat{M}_1, \hat{M}_2, \hat{M}_3$ et \hat{M}_4 , avec $\hat{M}_n = \frac{1}{\text{Card}(S^N)} \sum_{s \in S^N} y_s^n$ pour $n = 1, \dots, 4$.

D'autre part, soit une solution (f_1, \dots, f_K) (parmi les $j_1 \dots j_K$ solutions possibles), avec les probabilités a priori correspondantes (p_1, \dots, p_K) . Pour $1 \leq i \leq K$ et $1 \leq n \leq 4$, soit μ_{in} le $n^{\text{ième}}$ moment de la distribution f_i . L'adéquation entre l'image observée et la densité estimée $p_1 f_1 + \dots + p_K f_K$ est évaluée par $T = \left| \sum_{n=1}^4 [\hat{M}_n - (\sum_{i=1}^K p_i \mu_{in})] \right|$, et on choisit, parmi les $j_1 \dots j_K$ possibilités, celle qui minimise T .

4.2. Étalonage de la méthode proposée

Nous présentons ci-après quatre séries de résultats en utilisant quatre images fortement bruitées. La recherche des mélanges se fait par les méthodes SEM 1 et SEM 2, et le choix final est effectué par le test de la sous-section précédente. Nous considérons deux familles admissibles : distributions normales et distributions exponentielles. Nous distinguerons ainsi trois types de mélanges, les mélanges normale-normale (N-N), les mélanges exponentielle-exponentielle (E-E) et les mélanges normale-exponentielle (N-E) qui aboutissent à la même estimation que les mélanges exponentielle-normale. Notons enfin que les cas présentés comme n'ayant « pas de solution » sont ceux pour lesquels l'algorithme d'estimation des paramètres n'a pas convergé.

Les différents résultats sont présentés sur les Figures 3 – 6. Les segmentations présentées sur la première ligne sont obtenues sur la base des paramètres estimés par SEM 1, et celles présentées sur la deuxième ligne sont obtenues sur la base des paramètres estimés par SEM 2.

Ces différents exemples font apparaître que même dans les cas de bruitage extrême (les deux distributions bruitantes de la Figure 6 ont mêmes moyennes et mêmes variances) la méthode de segmentation non supervisée proposée, utilisant les modèles évolutifs et l'algorithme d'estimation des mélanges généralisés, reste fiable. Bien entendu, de tel bruitages n'existent que très rarement dans la pratique ; cependant, l'intérêt pratique des résultats présentés est également d'attester de l'excellente robustesse des méthodes non supervisées de segmentation étudiées.

4.3. Lecture de cartons d'orgue de barbarie

4.3.1. Description d'un carton

Sur une demande de l'association Argentique 1, et afin d'illustrer la rapidité et la précision des algorithmes de segmentation non supervisée d'images développés sur les diarbres de Markov cachés, nous nous sommes intéressés à la mise au point d'un protocole de lecture d'un carton d'orgue de barbarie à 27 touches. Ce type de carton mesure 154 mm de large. La note la plus haute, qui est un *Ré7* est située à 5 mm du bord gauche du carton. Chaque note est constituée par un trou rectangulaire, de 3,5 mm de largeur, et dont la longueur est fonction de la durée de la note dans la partition originale, et de la vitesse choi-

sie pour le défilement standard du carton sur l'instrument. Deux notes voisines, dans la succession non chromatique des notes d'un tel orgue, sont séparées, sur le carton, par un écart de 2 mm, ce qui positionne la note la plus basse, un *Do4*, à 9 mm du bord droit du carton.

4.3.2. Protocole d'acquisition de l'image

Le carton est entraîné à une vitesse de défilement constante et filmé par une caméra grand angle, à une résolution de 320×320 pixels. On reconstitue, ensuite, à partir des portions non déformées des images de la séquence filmée, l'image que nous devons segmenter.

Dans le cadre de notre travail, la vitesse de défilement du carton est limitée par la résolution à laquelle nous travaillons, à cause de la nécessaire précision sur l'interprétation de la durée des notes. En tenant compte de la précision de lecture de l'orgue mécanique, nous pouvons, raisonnablement, travailler avec une marge d'erreur, longitudinale de 2 mm. En particulier, l'espace minimum entre deux répétitions d'une même note sera de 2 mm

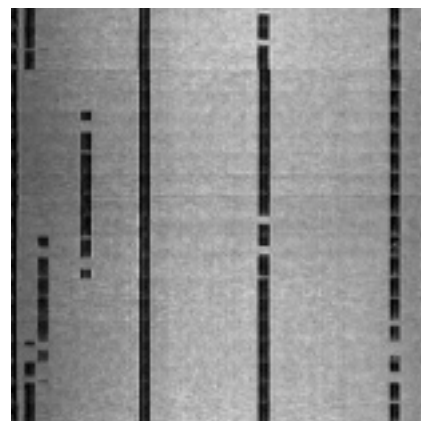


Figure 7. Image 1 « virtuosité ».

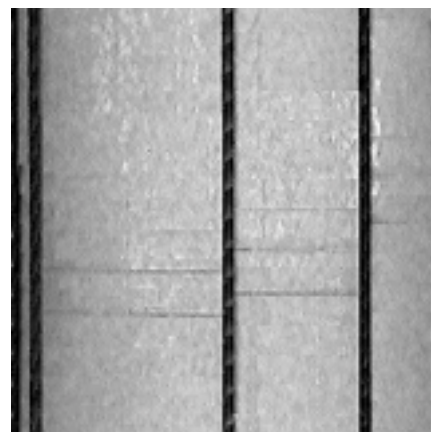


Figure 8. Image 2 « carton sale ».

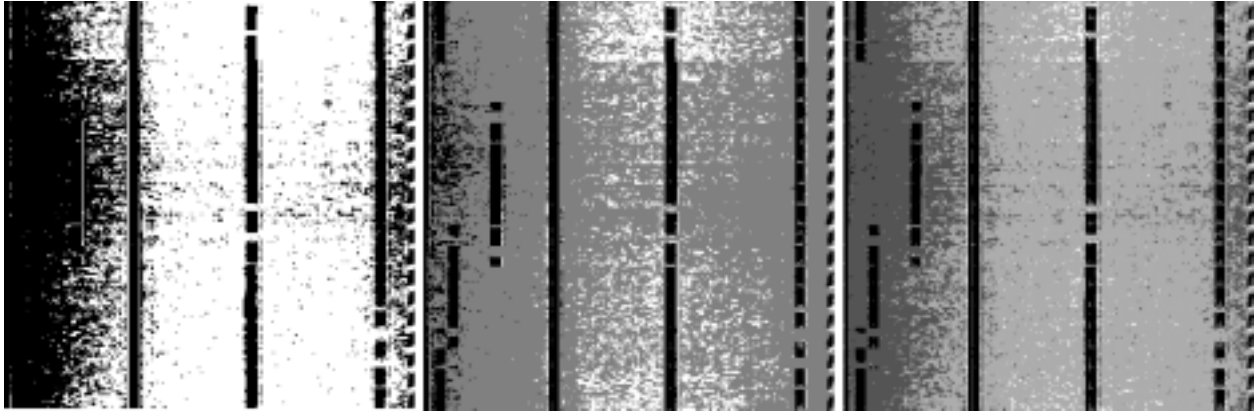


Figure 9. Classifications proportionnellement aux valeurs des pixels de l'image lue.

et la longueur minimale d'une note sera de 4 mm , la taille du poinçon permettant la perforation des cartons étant $3.5 \times 3.5\text{ mm}^2$. La vitesse de défilement du carton peut donc être fixée à 0.04 ms^{-1} , ce qui est un peu en dessous de la vitesse moyenne à laquelle le tourneur interprète son morceau. Nous ne pouvons cependant pas nous permettre de descendre sous la précision de segmentation longitudinale $1\text{ pixel} \rightarrow 1\text{ mm}$.

La largeur de l'image permet une plus grande précision latérale globale, de $1\text{ pixel} \rightarrow 0.5\text{ mm}$, qui compense la faible marge d'erreur longitudinale. Si l'on considère que la largeur théorique d'un trou correspond, alors, à 7 pixels , centrés sur une ligne dont la position est connue, parce que fixée par l'axe théorique de la note qu'ils représentent (standard du carton), il est raisonnable de considérer comme déterminante de l'une des deux classes, trou ou carton, la classe dominante, après segmentation, sur les 14 pixels formant l'unité de travail que nous nous sommes fixée. C'est ainsi que nous obtiendrons une segmentation interprétée du carton.

Les deux images auxquelles nous nous intéresserons sont présentées sur les Figures 7 et 8. La première recense toutes les difficultés musicales caractérisant les cartons d'orgue de barbarie, alors que la seconde soulève le problème des cartons anciens, tachés ou partiellement dégradés. L'étude de ces deux images nous permettra de mettre au point une méthode de lecture de cartons, grâce à laquelle nous serons en mesure de construire un fichier de type midi permettant l'écoute électronique du morceau ainsi que la copie de celui-ci sur un autre carton grâce à une interface appropriée.

4.3.3. Segmentation de l'image du carton 1

Les trois images de la Figure 9, déduites de l'image carton 1 et obtenues sans avoir recours à un algorithme de segmentation, montrent l'impossibilité d'exploiter directement l'information contenue dans le fichier de capture. À moins de perfectionner le dispositif d'acquisition des images, le recours à une méthode

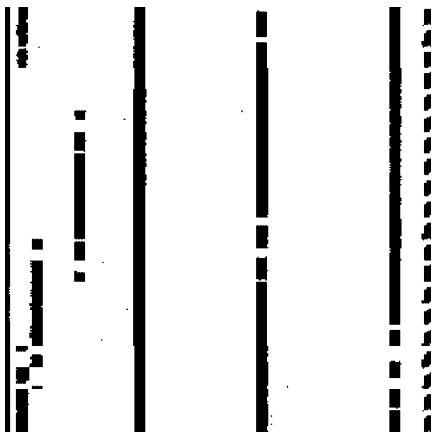


Figure 10. Segmentations N-N de l'image 1.

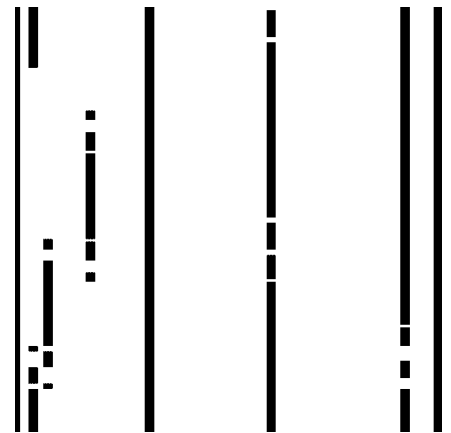


Figure 11. Image 1 interprétée.

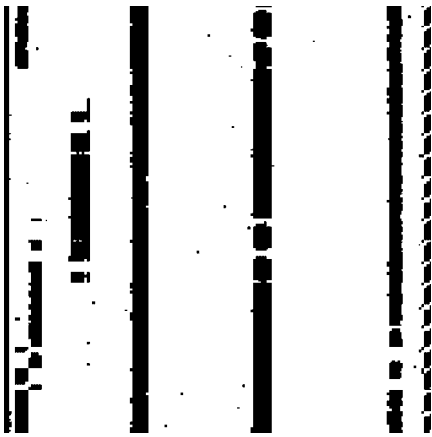


Figure 12. Segmentations N-E de l'image 1.

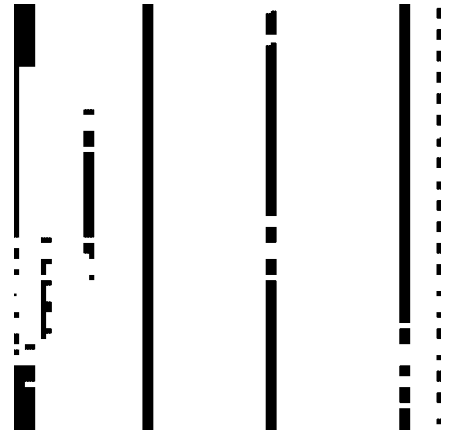


Figure 13. Segmentations E-E de l'image 1.

statistique de segmentation d'images apparaît comme une solution secourable.

Pour des raisons de rapidité de la segmentation, et dans un souci de qualité des résultats, il est nécessaire de morceler l'image originale à la résolution 320×320 , en 25 sous images à la résolution 64×64 . Cette solution permet de réaliser la segmentation de l'image de carton en moins de temps qu'il n'en faut pour acquérir l'image, ce qui équivaut à une interprétation du morceau en temps réel.

La segmentation sera menée selon deux classes, ce qui permettra une construction naturelle du fichier « lu », comme nous l'avons décrite plus haut. La première segmentation (cf. Figure 10) qui améliore incontestablement les résultats obtenus par la classification sans calculs, est effectuée dans le cadre d'un mélange de type N-N. L'image « interprétée » (cf. Figure 11) à partir de cette segmentation est très satisfaisante, aux approximations du protocole près.

Signalons aussi, qu'une segmentation, de chacune des 25 sous images, par la méthode de mélanges généralisés proposée précédemment, sélectionne, les 25 fois, le type de mélange N-N. Les segmentations dans les cas N-E, et E-E sont proposées pour information. La première (cf. Figure 12) est très proche de celle obtenue dans le cas N-N, à cela près que les classes « trou » sont, en général, plus larges. La connaissance théorique des zones de l'image susceptibles de porter la classe « trou », permet cependant une très bonne interprétation de cette image. La seconde (cf. Figure 13), au contraire, réduit la largeur de ces classes, qui apparaissent très nettement. Les travers de cette segmentation se révèlent notamment dans la partie gauche de l'image, où la restriction extrême des zones de classes « trou », fausse l'interprétation que nous pourrions en faire.

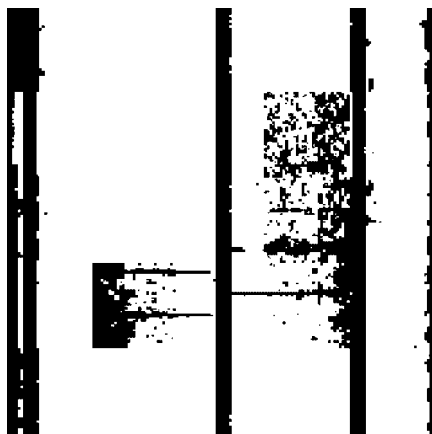


Figure 14. Segmentation N-N de l'image 2.



Figure 15. Segmentation E-E de l'image 2.

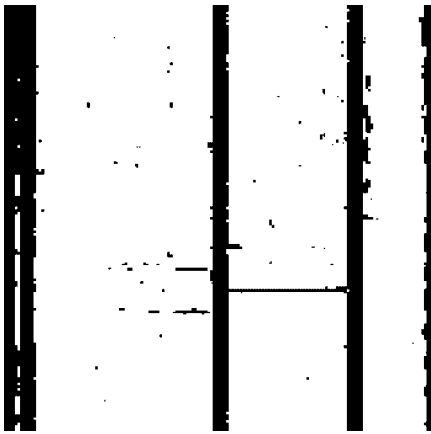


Figure 16. Segmentation N-E de l'image 2.

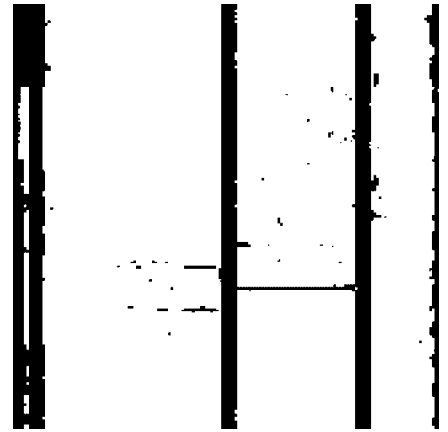


Figure 17. Image 2 segmentée par les mélanges généralisés.

4.3.4. Segmentation de l'image du carton sale

Le carton de la Figure 8 que nous essayons de lire ici, bien qu'il soit très simple d'un point de vue musical, soulève deux types de difficultés qui ne touchaient pas le premier carton. Il a d'une part, été taché par le vernis utilisé pour durcir les bordures des trous et éviter leur usure, et il a, d'autre part, été plié à plusieurs endroits. Ces problèmes se manifestent, au niveau de l'image à segmenter, par des taches anormalement claires pour le vernis, et par des traits sombres que l'on peut confondre avec les trous, pour les pliures.

La complexification du travail devient claire lorsque l'on tente de segmenter l'image carton sale par la méthode de modélisation gaussienne classique du bruit. L'image segmentée (cf. Figure 14) n'est pas exploitable pour une interprétation musicale correcte du carton.

Il apparaît, en fait, que si la segmentation sur le modèle E-E (cf. Figure 15) n'est pas plus pertinente, il n'en est pas de même de celle effectuée sur le modèle N-E présentée sur la Figure 16. Cette dernière segmentation, bien que globalement moins précise que la segmentation N-N en de nombreuses zones, identifie les trois notes et n'entraîne pas l'apparition artificielle de fausses notes. Remarquons également que la méthode d'estimation fondée sur les mélanges généralisés privilégie encore les segmentations N-N et ne choisira pas toujours la segmentation N-E même lorsqu'elle est pertinente.

4.3.5. Une méthode de lecture

Le problème qui se pose dans ce cas très particulier de segmentation d'images réelles, typiquement adapté au cas des mélanges généralisés, est que le critère de choix général que nous avons décrit dans le chapitre précédent mène parfois à un choix non optimal de segmentation. Il convient donc de définir un critère de choix pour la méthode des mélanges généralisés, spécifiquement adapté au problème qui nous intéresse. L'interprétation du

carton demande, en effet, une grande précision dans la segmentation, que l'on peut atteindre grâce à une bonne connaissance des contraintes *a priori*. Le nouveau critère consiste à choisir, pour chacune des 25 sous-images, celle des deux segmentations issues des estimations pour des mélanges de type N-N et N-E, qui aura le moins souvent attribué la classe « trou », à des variables que nous savons nécessairement classées « carton » (puisque se situant dans l'intervalle de 2mm séparant latéralement deux notes différentes du carton).

La nouvelle méthode d'estimation des paramètres de mélanges généralisés que nous proposons, définie pour la segmentation d'images de cartons d'orgue de barbarie, augmente la durée totale de la segmentation menée sans les mélanges généralisés, mais se révèle nécessaire. L'augmentation de la durée des calculs permet cependant toujours la lecture en temps réel du carton. Cette procédure, appliquée à l'image carton sale, permet effectivement de corriger, par le choix de la segmentation issue

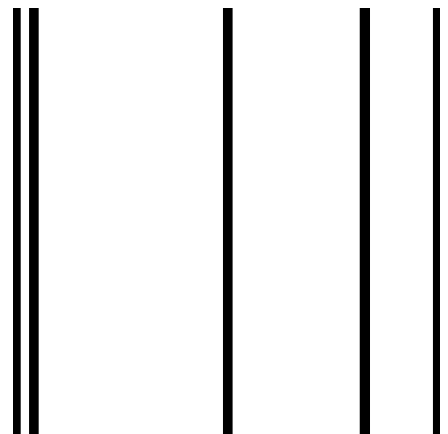


Figure 18. Image 2 interprétée.

du type N-E, les quatre sous images dégradant précédemment le résultat de la segmentation N-N. L'image ainsi obtenue optimise la segmentation et par conséquent, la qualité de la lecture du carton (cf. Figure 17). Notons, pour conclure, que cette méthode, appliquée à l'image carton 1 permet d'aboutir, conformément à ce qu'on pouvait prévoir, à l'image interprétée recherchée (cf. Figure 18).

5. Conclusions

Nous avons traité dans cet article du problème de l'utilisation des arbres de Markov cachés (AMC) à des fins de segmentation non supervisée d'images. De manière générale, de tels modèles ont été utilisés avec succès dans de nombreuses situations. Citons l'imagerie bathymétrique [4], astronomique [5], ou satellitaire [11]. Plus généralement, ce type de modèles est très bien adapté à l'imagerie multi-résolution et l'on trouvera dans [20] une riche bibliographie concernant le sujet. Dans ce contexte, nous nous sommes intéressés, au plan méthodologique, à trois points :

(i) les modèles classiques, dans lesquels les transitions ne dépendent pas de la résolution, présentent souvent des effets indésirables, dits « effets de blocs ». Nous avons proposé des modèles dits « évolutifs », dans lesquels la probabilité de conservation des classes croît lorsque l'on se rapproche de l'échelle la plus fine. Nous avons constaté, à partir des différentes expérimentations, l'intérêt des modèles évolutifs dans l'atténuation desdits effets (voir aussi [15]) ;

(ii) il existe un grand nombre de méthodes d'estimation de paramètres et nous avons testé, de manière relativement complète, trois d'entre elles : « Expectation – Maximization » (EM [13, 14]), « Stochastic » EM (SEM [3], et « Iterative Conditional Estimation » (ICE [8-10]). Globalement, le comportement de ces méthodes est comparable avec, cependant, un léger avantage pour le SEM. Notons que cette conclusion rejoint celles d'une étude comparable effectuée dans le cadre des chaînes de Markov cachées présentée dans [3], où ces trois méthodes ont également montré des comportements très proches. De plus, elles restent efficaces jusqu'aux niveaux très élevés du bruit. Cette propriété est intéressante dans la mesure où elle implique que la dégradation de l'efficacité d'une méthode non supervisée de segmentation liée à l'augmentation du niveau de bruit n'est pas due à la phase de l'estimation des paramètres. Cela montre qu'il reste « une marge de manœuvre » pour utiliser de telles méthodes dans des conditions de bruitages plus complexes. Comme cela a été le cas pour les chaînes [10] et les champs [8] de Markov cachés, cette constatation nous a conduit à considérer le problème de l'estimation des mélanges généralisés ;

(iii) une nouvelle méthode d'estimation de mélanges généralisés, ces derniers étant considérés dans le cadre des AMC, a ensuite été proposée. Le problème de mélange généralisé survient lorsque la forme même du bruit relatif à chaque classe n'est pas connue et peut varier avec cette dernière. La méthode

proposée, testée dans le cadre des mélanges faisant intervenir des lois gaussiennes et exponentielles, s'avère efficace jusqu'aux niveaux très élevés du bruit.

La dimension supplémentaire induite par le nouveau modèle incluant les mélanges généralisés est également apparue à travers l'étude menée pour définir une méthode de lecture des cartons d'orgue de barbarie. En effet, la méthode proposée, testée sur plusieurs cartons complets, donne de très bons résultats. Les retouches musicales à apporter sur le fichier interprété sont raisonnablement peu fréquentes et généralement concentrées sur deux ou trois mesures pour lesquelles les contraintes de virtuosité du morceau, ou la qualité locale de l'acquisition de l'image perturbent la classification. Les principales limites de cette technique de lecture de cartons d'orgue de barbarie sont celles fixées par le protocole d'acquisition de l'image. Une résolution de capture supérieure (sans perte concomitante de rapidité), permettrait sans doute de gagner en fiabilité. Une exploitation, en parallèle, des différents canaux RVB d'une capture couleur, optimisée par la possibilité de colorer le support qui caractérise la couleur de la classe « trou », est actuellement en cours de développement.

Comme perspective nous pouvons envisager l'extension des diverses méthodes proposées dans cet article aux arbres de Markov couple, modèles qui généralisent strictement les AMC et dont les premières applications dans le contexte gaussien semblent prometteuses [16].

Références

- [1] B. BENMILOUD, W. PIECZYNSKI, *Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images*, Traitement du Signal, Vol. 12, No. 5, pp. 433-454, 1995.
- [2] C. BOUMAN, M. SHAPIRO M., *A multiscale random field model for bayesian image segmentation*, IEEE Transactions On Image Processing, Vol 3, No. 2, 162-177, 1994.
- [3] G. CELEUX, J. DIEBOLT, The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem, *Comput. Statist. Quarter.* 2, pp. 73-82, 1985.
- [4] C. COLLET, J.-N. PROVOST, P. ROSTAING, P. PÉREZ, P. BOUTHEMY, *Segmentation bathymétrique d'images multispectrales SPOT*, Traitement du Signal, Vol. 18, No. 1, pp. 1-14, 2001.
- [5] C. COLLET, M. LOUYS, J.-N. PROVOST, A. OBERTO, *Fusion of astronomical multiband images on a Markovian quadree*, Proceedings of the Fifth International Conference on Information Fusion, 8-11 July 2002, Annapolis MD, USA, Vol. 1, pp. 396 – 403, 2002.
- [6] C. COLLET, F. MURTAGH, *Multiband segmentation based on a hierarchical Markov model*, Pattern Recognition, to appear, 2004.
- [7] M. CROUSE, R. NOWAK, R. BARANIUK, Wavelet-based statistical signal processing using hidden Markov models, *IEEE Trans.Signal Processing*, Vol. 46, pp. 886-902, 1998.
- [8] Y. DELIGNON, A. MARZOUKI, W. PIECZYNSKI, *Estimation of generalized Mixture and its application in Image Segmentation*, IEEE Trans. on Image Processing, Vol. 6, No. 10, 1764-1774, 1997.

- [9] J.-P. DELMAS, *An equivalence of the EM and ICE algorithm for exponential family*, IEEE Trans. on Signal Processing, Vol. 45, No. 10, pp. 2613-2615, 1997.
- [10] N. GIORDANA, W. PIECZYNSKI, *Estimation of Generalized multi-sensor Hidden Markov Chains and Unsupervised Image Segmentation*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 5, 465-476, 1997.
- [11] L. HUBERT-MOY, A. COTONNEC, L. LE DU, A. CHARDIN, P. PÉREZ, A comparison of parametric classification procedures of remotely sensed data applied on different landscape units, Remote Sensing of Environment, Vol. 75, N° 2, pp. 174-187, 2001.
- [12] J.-M. LAFERTÉ, *Contribution à l'analyse d'images par modèles markoviens sur des graphes hiérarchiques – Application à la fusion de données multirésolution*, Thèse de doctorat de l'Université de Rennes I, 1996.
- [13] J.-M. LAFERTÉ, P. PEREZ, F. HEITZ, *Discrete Markov image modeling and inference on the quad-tree*, IEEE Trans. on Image Processing, Vol. 9, No 3, 390-404, 2000.
- [14] J. MCLACHLAN, T. KRISHNAN, *EM Algorithm and Extensions*, Wiley, 1997.
- [15] E. MONFRINI, *Identifiabilité et Méthode des Moments dans les mélanges généralisés de distributions du système de Pearson*, Thèse de l'Université Paris VI, soutenue à l'Université Paris VI le 4 janvier 2002.
- [16] E. MONFRINI, J. LECOMTE, F. DESBOUVRIES, W. PIECZYNSKI, (2003). *Image and Signal Restoration using Pairwise Markov Trees*, IEEE Workshop on Statistical Signal Processing (SSP 2003), Saint Louis, Missouri, USA, September, 28-October 1, 2003.
- [17] J.-N. PROVOST, C. COLLET, P. ROSTAING, P. PÉREZ, P. BOUTHEMY, *Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps*, Computer Vision and Image Understanding, Vol. 93, No. 2, pp. 155-174, 2004.
- [18] F. SALZENSTEIN, W. PIECZYNSKI, *Sur le choix de méthodes de segmentation statistique d'images*, Traitement du signal, Vol 15, No. 2, 119-128, 1998.
- [19] C. K. L. WILLIAMS, X. FENG, *Tree-Structured belief networks as models of images*, Proceedings of ICANN'99, 1999.
- [20] A. S. WILLSKY, *Multiresolution Markov models for signal and image processing*, Proceedings of IEEE, Vol. 90, No. 8, pp. 1396-1458, 2002.



Wojciech **Pieczynski**

Wojciech Pieczynski a reçu un doctorat d'État en Statistique Mathématique de l'Université Paris VI en 1986. Après avoir enseigné au Centre Universitaire de Tiaret, Algérie (1981-83), à l'Université de Brazzaville, Congo (1983-87), et à l'École Nationale Supérieure des Télécommunications de Bretagne (1987-90), il est professeur à l'INT depuis 1990, et responsable du département Communications, Images, et Traitement de l'Information (CITI) depuis septembre 2004. Ses recherches portent sur les modèles de Markov et les traitements statistiques d'images.



Emmanuel **Monfrini**

Emmanuel Monfrini a obtenu, en 2002, un doctorat de Mathématiques de l'Université Paris VI. Membre du Laboratoire de Statistique Théorique et Appliquée, il poursuit des travaux de recherche autour du thème central des distributions de mélanges finis généralisés de lois de probabilité. Collaborant avec l'INT depuis la préparation de sa thèse, il étudie les processus stochastiques et contribue au développement de méthodes statistiques de segmentation d'images.

